

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1995	1 MAY 92 To final technical report 30 Apr 95
4. TITLE AND SUBTITLE		5. FUNDING NUMBERS
A self-organizing neural network architecture for auditory and speech perception with applications to acoustic and other temporal prediction problems		F49620-92-J-0225  61102F  23B/AS
6. AUTHOR(S)		
Stephen Grossberg		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)		8. PERFORMING ORGANIZATION REPORT NUMBER
Center for Adaptive Systems and Department of Cognitive and Neural Systems Boston University, Boston, MA 02215		AFOSR-TR-95  0498
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		
Air Force Office of Scientific Research/NL Life Sciences Directorate Bolling AFB, DC 20332-6448		
Dr John F. Tangney		
11. SUPPLEMENTARY NOTES		
12a. DISTRIBUTION/AVAILABILITY STATEMENT		12b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.		S DTIC ELECTED AUG 30 1995 D  F
13. ABSTRACT (Maximum 200 words)		
<p>This project is developing autonomous neural network models for the real-time perception and production of acoustic and speech signals. The models have disclosed a common mechanism of nonlinear resonance that attentively reorganizes and groups acoustic data while suppressing unexpected noise. The SPINET pitch model transforms acoustic input into a spatial map of pitch whose properties simulate the key pitch data. SPINET was embedded into an ARTSTREAM model for auditory scene analysis that separates multiple sound sources from each other. The model groups frequency components based on pitch and spatial location cues into different streams. The model simulates psychophysical grouping data, such as frequency grouping across noise or ear of origin. These resonant streams input to an ARTPHONE model for variable-rate speech categorization. Computer simulations quantitatively generate experimentally observed category boundary shifts for VC-CV pairs, including why the interval to hear a double (VC<sub>1</sub>-C<sub>1</sub>V) stop is 150 msec longer than that to hear two different stops (VC<sub>1</sub>-C<sub>2</sub>V). This model uses resonant feedback between list categories and an automatically gain-controlled working memory.</p>		

14. SUBJECT TERMS		15. NUMBER OF PAGES
		31 pages
DTIC QUALITY INSPECTED 5		16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT
unclassified	unclassified	u
		20. LIMITATION OF ABSTRACT

PUBLICATIONS PARTIALLY SUPPORTED BY  
THE AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

CONTRACT AFOSR F49620-92-J-0225

MAY 1, 1992—APRIL 30, 1995

CENTER FOR ADAPTIVE SYSTEMS  
AND  
DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS  
BOSTON UNIVERSITY

1. Boardman, I.S. (1994). Neural network models of temporal processing in speech perception and motor control. PhD dissertation, Department of Cognitive and Neural Systems, Boston University.
2. Boardman, I., Cohen, M.A., and Grossberg, S. (1993). Variable rate working memories for phonetic categorization and invariant speech perception. In **Proceedings of the world congress on neural networks**, Portland, **III**, 2–5. Hillsdale, NJ: Erlbaum Associates. (%&)
3. Boardman, I., Grossberg, S., and Cohen, M.A. (1994). Neural dynamics of phonetic trading relations for variable-rate CV syllables. **Technical Report CAS/CNS-TR-94-037**, Boston University. Submitted for publication. (%)
4. Bradski, G., Carpenter, G.A., and Grossberg, S. (1994). STORE working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics*, **71**, 469–480. (#%+&)
5. Bradski, G. and Cohen, M.A. (1993). A preliminary look at a fast learning architecture for speaker independent speech recognition. In **Proceedings of the world congress on neural networks**, Portland, **III**, 37–39. Hillsdale, NJ: Erlbaum Associates. (+@)
6. Bullock, D., Fiala, J.C., and Grossberg, S. (1994). A neural model of timed response learning in the cerebellum. *Neural Networks*, **7**, 1101–1114. (%+&)
7. Bullock, D., Grossberg, S., and Mannes, C. (1993). The VITEWRITE model of handwriting production. In **Proceedings of the world congress on neural networks**, Portland, **I**, 507–511. Hillsdale, NJ: Erlbaum Associates. (%+&)
8. Carpenter, G.A. and Govindarajan, K.K. (1993). Evaluation of speaker normalization methods for vowel recognition using fuzzy ARTMAP and K-NN. In **Proceedings of the world congress on neural networks**, Portland, **III**, 10–15. Hillsdale, NJ: Erlbaum Associates. (#%+&)
9. Carpenter, G.A. and Govindarajan, K.K. (1993). Speaker normalization methods for vowel recognition: Comparative analysis using neural network and nearest neighbor classifiers. **Technical Report CAS/CNS-TR-93-039**, Boston University. Submitted for publication. (%#+@)
10. Carpenter, G.A. and Grossberg, S. (1994). Integrating symbolic and neural processing in a self-organizing architecture for pattern recognition and prediction. In V. Honavar and L. Uhr (Eds.), **Artificial intelligence and neural networks: Steps toward principled integration**. San Diego: Academic Press, pp. 387–421. (#%+&)
11. Carpenter, G.A. and Grossberg, S. (1995). A neural network architecture for autonomous learning, recognition, and prediction in a nonstationary world. In S.F. Zornetzer, J.L. Davis, and C. Lau (Eds.), **An introduction to neural and electronic networks: Second edition**. New York: Academic Press, pp. 465–482. (%#+&)

19950828 003

12. Carpenter, G.A. and Grossberg, S. (1995). Adaptive resonance theory. **Technical Report CAS/CNS-TR-94-034**, Boston University. In M.A. Arbib (Ed.), **Handbook of brain theory and neural networks**. Cambridge, MA: MIT Press, in press. (%&)
13. Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1995). A fuzzy ARTMAP nonparametric probability estimator for nonstationary pattern recognition problems. **Technical Report CAS/CNS-TR-93-047**, Boston University. *IEEE Transactions on Neural Networks*, in press. (\*%#+&)
14. Cohen, M.A. (1993). Neural network models of speech and language perception and recognition. In **Proceedings of the world congress on neural networks**, Portland, **III**, 1. Hillsdale, NJ: Erlbaum Associates.
15. Cohen, M.A. and Govindarajan, K.K. (1994). Influence of silent interval distribution on stop consonant clusters. In preparation. (#+)
16. Cohen, M.A. and Grossberg, S. (1993). Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants. **Technical Report CAS/CNS-TR-93-051**, Boston University. Submitted for publication. (%@)
17. Cohen, M.A., Grossberg, S., and Priebe, C. (1993). Neural control of interlimb coordination and gait timing in bipeds and quadrupeds. **Technical Report CAS/CNS-TR-93-004**, Boston University. Submitted for publication. (@%&)
18. Cohen, M.A., Grossberg, S., and Priebe, C. (1993). A neural pattern generator that exhibits arousal-dependent human gait transitions. In **Proceedings of the world congress on neural networks**, Portland, **IV**, 285–288. Hillsdale, NJ: Erlbaum Associates. (\*%+&)
19. Cohen, M.A., Grossberg, S., and Priebe, C. (1993). Frequency-dependent phase transitions in the coordination of human bimanual tasks. In **Proceedings of the world congress on neural networks**, Portland, **IV**, 491–494. Hillsdale, NJ: Erlbaum Associates. (\*%+&)
20. Cohen, M.A., Grossberg, S., and Priebe, C. (1993). Quadruped gait transitions from a neural pattern generator with arousal modulated interactions. In **Proceedings of the world congress on neural networks**, Portland, **II**, 610–613. Hillsdale, NJ: Erlbaum Associates. (\*%+&)
21. Cohen, M.A., Grossberg, S., and Wyse, L. (1995). A spectral network model of pitch perception. **Technical Report CAS/CNS-TR-92-024**, Boston University. *Journal of the Acoustical Society of America*, in press. (%&)
22. Cohen, M.A. and Priebe, C. (1993). A biomechanical model of human oculomotor plant kinematics based upon geometric algebra. **Technical Report CAS/CNS-TR-93-044**, Boston University. Submitted for publication. (&)
23. Govindarajan, K.K. (1994). Invariant speech recognition and auditory object formation: Neural models and psychophysics. PhD dissertation, Department of Cognitive and Neural Systems, Boston University. (#+)
24. Govindarajan, K.K. and Cohen, M.A. (1994). Influence of silence duration in the perception of stop consonant clusters. *Journal of the Acoustical Society of America*, **95**, 2978. Poster presentation at the 127th annual meeting of the Acoustical Society, Cambridge, MA, 6–10 June. (#+)
25. Govindarajan, K.K., Grossberg, S., Wyse, L.L., and Cohen, M.A. (1994). A neural network model of auditory scene analysis and source segregation. **Technical Report CAS/CNS-TR-94-039**, Boston University. Submitted for publication. (%#+)

26. Grossberg, S. (1993). Self-organizing neural networks for stable control of autonomous behavior in a changing world. In J.G. Taylor (Ed.), **Mathematical approaches to neural networks**. Amsterdam: Elsevier Science Publishers, pp. 139–197. (%&)
27. Grossberg, S. (1995). The attentive brain. Invited article for *American Scientist*, in press. (%&)
28. Grossberg, S., Boardman, I., and Cohen, M.A. (1994). Neural dynamics of variable-rate speech categorization. **Technical Report CAS/CNS-TR-94-038**, Boston University. Submitted for publication. (%)
29. Grossberg, S. and Grunewald, A. (1993). Statistical properties of single and competing nonlinear fast-slow oscillations in noise. In **Proceedings of the world congress on neural networks**, Portland, **IV**, 303–307. Hillsdale, NJ: Erlbaum Associates. (%+&)
30. Grossberg, S. and Grunewald, A. (1994). Spatial pooling and perceptual framing by synchronizing cortical dynamics. In M. Marinaro and P.G. Morasso (Eds.), **Proceedings of the international conference on artificial neural networks (ICANN-94)**. New York: Springer-Verlag, pp. 10–15. (%&)
31. Grossberg, S. and Grunewald, A. (1994). Synchronized neural activities: A mechanism for perceptual framing. In **Proceedings of the world congress on neural networks**, San Diego, **IV**, 655–660. Hillsdale, NJ: Erlbaum Associates. (%&)
32. Grossberg, S. and Grunewald, A. (1994). Cortical synchronization and perceptual framing. **Technical Report CAS/CNS-TR-94-25**, Boston University. Submitted for publication. (%&)
33. Grossberg, S. and Merrill, J.W.L. (1995). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. **Technical Report CAS/CNS-TR-93-065**, Boston University. *Journal of Cognitive Neuroscience*, in press. (%&)
34. Grossberg, S., Mingolla, E., and Williamson, J. (1995). Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. **Technical Report CAS/CNS-TR-94-001**, Boston University. *Neural Networks* (Special Issue on Automatic Target Recognition), in press. (%+&)
35. Grossberg, S., Mingolla, E., and Williamson, J. (1995). A multiple scale neural system for boundary and surface representation of SAR data. In **Proceedings of the IEEE workshop on neural networks for signal processing**. New York: IEEE Publishing Services, in press. (%+&)
36. Grossberg, S., Priebe, C., and Cohen, M.A. (1994). Neural control of interlimb oscillations, I: Human bimanual coordination. **Technical Report CAS/CNS-TR-94-021**, Boston University. Submitted for publication. (@%+&)
37. Grunewald, A. (1994). A neuron model with variable ion concentrations. In **Proceedings of the world congress on neural networks**, San Diego, **IV**, 368–372. Hillsdale, NJ: Erlbaum Associates.
38. Grunewald, A. and Grossberg, S. (1994). Binding of object representations by synchronous cortical dynamics explains temporal order and spatial pooling data. In A. Ram and K. Eiselt (Eds.), **Proceedings of the sixteenth annual conference of the Cognitive Science Society**. Hillsdale, NJ: Erlbaum Associates, pp. 387–391. (%&)
39. Kincaid, T.G. and Cohen, M.A. (1994). Sufficient conditions for WTA behavior of a shunting inhibition network. Submitted for publication.
40. Priebe, C. (1994). Neural dynamics of gaze and gait: Spatial sensory-motor control and temporal pattern generation. PhD Dissertation, Department of Cognitive and Neural Systems, Boston University.
41. Priebe, C., Grossberg, S., and Cohen, M.A. (1994). Neural control of interlimb oscillations, II: Biped and quadruped gaits and bifurcations. **Technical Report CAS/CNS-TR-94-022**, Boston University. Submitted for publication. (@%+&)

- \* Also supported in part by the AFOSR URI [expired].
- @ Also supported in part by the Army Research Office [expired].
- % Also supported in part by the Advanced Research Projects Agency [expired].
- # Also supported in part by British Petroleum [expired].
- + Also supported in part by the National Science Foundation.
- & Also supported in part by the Office of Naval Research.

Accesion For	
NTIS	CRA&I
DTIC	TAB
Unannounced	
Justification	
By _____	
Distribution / _____	
Availability Codes	
Dist	Avail and/or Special
A-1	

## RESEARCH SUMMARIES

### Pitch Perception

One of the major human auditory abilities is the effortless identification of multiple speakers while simultaneously understanding the content of their speech. There is considerable evidence that an important cue for the segregation of speech by speaker is the pitch of the utterance. Cohen, Grossberg, and Wyse have developed a neural model of pitch perception, called the SPINET model, and this work has been accepted for publication in the *Journal of the Acoustical Society of America*. A front end for the model has been developed to take real time acoustic input and operate on auditory signals. The prior model used idealized pitch representations as input. The revised model therefore has the capacity to operate in real time on the speech data. This model provides a unified explanation of the key experimental data of pitch perception and is part of our rapidly developing theory of auditory object recognition. The model computes a spatial representation of pitch as one of several steps in the process of auditory source segregation and streaming. Simulated data include pitch perception with mistuned or missing components, shifted harmonics, and various types of noise; the auditory dominance region; octave shifts to ambiguous stimuli; Shepard auditory barberpole illusions; and the Ritsma and Engle (1964) quasi FM pitch data. The model is a spectral model which can nonetheless simulate data commonly thought to have a temporal base. The model provides a basis for linking featural and spatial information to rapidly locate and bind together the auditory signals that correspond to a target in space. Research on such target localization has begun. [Article 20]

### Design of STORE Working Memories for Invariant Storage and Learning of Auditory Sequences

The work on this project begins to model the interaction between a working memory that temporarily stores speech tokens and its feedback interactions with a speech categorization network. This working memory has the critical property that the pattern of stored information remains invariant under variable speaking rates. In addition, the project showed how a working memory could be designed such that all possible groupings of stored items could be learned in a stable fashion, even if a grouping was later stored and learned as part of a larger grouping. Thus the word MY is not forgotten when it is later learned as part of the word MYSELF. This was shown for arbitrary temporal streams of data, occurring with essentially arbitrary rates, delays, and repetitions. A surprising property of such a working memory is that it can produce a primacy, recency, or bowed temporal order gradient, as are known to occur in a number of language paradigms. [Article 4]

### Application of the STORE Working Memory Design and Peripheral Filtering to Speech Recognition

One competence routinely displayed by humans but often neglected in approaches to machine recognition is the modification of prior auditory representations by subsequent auditory signals. This modification has both a short term and long term aspect. By short term, a time scale of a couple of syllables is meant; by long term, the time scale of a few seconds to a few minutes of discourse. As an example of the first phenomenon, it is known that the duration of a subsequent vowel can shift the category boundary between recognition of the prior consonant between stop and glide. It was found that, as in the model of early speech filtering discussed below [Article 16], sustained and transient detectors which measure the duration of the speech segments and the slope of transitions can be used to model short term modification. The ratio of the outputs of these sustained and transient detectors has been successfully used to model the shift in the boundary between a (/b/) and (/w/) as a function of speech duration. [Article 3].

In additional work, we have also studied the long term modification of prior auditory context. Repp (1980) has shown the long term statistical distribution of silence between two stop stimuli shifts and the category boundary between hearing a single (/aga/) or cluster of consonants (/agpa/) cluster. The same statistical effect holds for the category boundary between single (/aga/) and geminant (/agga/) consonant clusters. However, the mean length of the category boundary for this distinction is far longer (200 milliseconds) than for the single cluster distinction (70 milliseconds). Repp determined this by graphing the probability of hearing one versus two stops as a function of the silent interval separating the two stops. This psychometric functions shifted depending upon the statistics of the stimulus presentations.

This work, carried out collaboratively with our student Ian Boardman, has modified the the STORE architecture so as to generate a representation of stored speech tokens that is independent of the speech rate. Furthermore, the psychometric function for the single/geminant and single/cluster conditions was fit [Article 2]. However, this model suffered from two deficiencies. First, there was no categorical storage in this variant in the model, so that groupings of speech units could not be learned. Second, there was no intrinsic difference between the adaptation to repetitions of the same consonant (the single geminant distinction) and of two differing consonants (the single cluster distinction). As a result, the difference in timing between single cluster and single geminant could not be explained.

This explanatory gap was overcome by embedding the invariant STORE working memory into a larger architecture wherein signals from the working memory input to a chunking network that groups speech tokens into large language units. Feedback from the chunking network selects consistent speech groupings and inhibits inconsistent groupings, as occurs during phonemic restoration. Resonances bewteen mutually consistent working memory and chunking representations constitute the speech code. Using this expanded model of speech perception, the Repp (1980) psychometric data was quantitatively fit [Article 28].

The comparison between model response and data led to further empirical questions concerning the original experimentation. Repp (1980) averaged his psychometric functions across multiple subjects. The averaged data showed a shallower psychometric function when the inputs had a broader temporal distribution. It was difficult to fit this data with the model. Furthermore, the model stored variables which were sensitive to the statistics of stimulus presentations but were relatively insensitive to short term sequential effects in presentation of the stimulus interval.

### **Experimental Tests of the Phonetic Network**

To elucidate these questions, further experiments were done by Govarindarajan and Cohen [Article 24]. The experiments of Repp (1980) were replicated using natural speech signals. However, the individual subject data were kept. Furthermore, the statistical distributions of presentation were varied while fixing the mean to study whether individual subjects were sensitive to the mean or other parameters of the distribution. Finally the subject responses were studied conditional to the prior distribution of the preceding two trials while fixing the mean of the prior distribution. In this way, short term sequential effects of the immediately prior data could be studied independent of the mean presentation interval.

In support of our model, the individual data had relatively steep psychometric functions in all conditions, but when the stimuli were presented with a broad temporal interval, the thresholds for detecting one or two stimuli varied considerably across the population. This strongly suggests that the shallow psychometric function was not a function of individual subject but was a property of the averaging process. Small but systematic sequential effects will be analysed using the model reported. [Article 28].

### **Speech Filtering and Coarticulation**

The rate of reception of normal spoken language places severe demands on the sensory, motor, and cognitive mechanisms which control the speech production process. The vocal

musculature often cannot keep up with the rate of transmission of spoken speech. Nature's solution to the motoric problem is to overlap significant parts of gestural motion for adjacent speech stops and consonants, a phenomenon called coarticulation.

This coarticulation creates corresponding perceptual problems, because the acoustic stream for the same phonetic percept is of necessity context-dependent. An early stage of auditory neural processing must be able to separate and represent these segments in a context-independent fashion to disambiguate these coarticulations.

Parallel processing streams that are sensitive to sustained and transient features are suggested to play such a role in speech perception. Analogous parallel streams are known to occur in vision. We have submitted a paper for publication [Article 16] which models these auditory sustained and transient mechanisms and shows how their combination can disambiguate sonorant (vowels [/i/], nasal [m,n], and glides [y,w]) from non-sonorant stops /t/ and fricatives /s,z/) phonetic classes. Furthermore, these detectors can distinguish between the major subcategories of these classes. This work hereby proposes a partial description of how the human auditory apparatus solves the perceptual problem caused by coarticulation. Later processing is greatly simplified by this processing stage.

### Auditory Scene Analysis: Streaming and Segregation

The ability of a listener to pay attention to a particular speaker in a noisy room or in a room with other speakers, e. g. at a cocktail party, attests to the robustness of the auditory peripheral system. Even though these multiple sound sources mix together their harmonics to produce one signal at the listener's ear, the auditory system is capable of teasing apart this jumbled signal to recognize different mental objects for different sound sources. The ability to segregate these different signals has been termed auditory scene analysis (Bregman, 1990). The scene analysis corresponds to the mechanisms by which the auditory system selectively groups certain acoustic features while excluding others to form internal representations of auditory objects.

These human competencies suggest important problems for artificial speech recognition in natural environments. A key problem for artificial speech recognition is the effortless simultaneous segregation of naturally occurring sounds into auditory objects to be attended to separately and background noise to be ignored where possible. Recent collaborative modeling with our student Krishna Govindarajan [Article 25] has been used to simulate seminal data discussed in the literature for auditory object formation. Our simulations start from the raw signals used in experimental manipulations and reliably selects a different neural stream in real time for each auditory object heard. This data includes: segregation such as functional segregation of anharmonic simple tones into separate streams (Moore *et al.*, 1985), the continuation of tones into noise (Miller and Licklider, 1950), the bounce percept obtained when two linear ramping FM glides cross (Halperin, 1977; Tougas and Bregman, 1990), and the Steiger (1990) diamond shaped FM stimulus. The structure of this model also clarifies how spatial localization could be used to segregate auditory objects.

The streaming model uses the SPINET pitch model as a front end. It shows how multiple representations of spectral streams interact reciprocally with multiple representations of pitch streams through cooperative and competitive feedback to automatically segregate appropriate groupings of auditory spectral components into one stream, while preventing these components from being used by another stream. Remarkably, a key circuit in insuring this selectivity is adapted from Adaptive Resonance Theory, or ART (Carpenter and Grossberg, 1991, 1993; Grossberg, 1980). Bottom-up and top-down reciprocal exchanges set up a resonance between consistent spectral groupings and pitches, while suppressing inconsistent spectral components, within each stream. These ART mechanisms are computationally homologous to the ones used to model the working memory and chunking feedback network that simulated the Repp (1980) data on category boundaries. Thus some general principles of auditory signal processing seem to be coming into view.

## **Neural Dynamics of Phonetic Trading Relations for Variable-Rate CV Syllables**

The perception of CV syllables exhibits a trading relationship between voice onset time (VOT) of a consonant and duration of a vowel. Percepts of [ba] and [wa] can, for example, depend on the durations of the consonant and vowel segments, with an increase in the duration of the subsequent vowel switching the percept of the preceding consonant from [w] to [b]. A neural model, called PHONET, is proposed to account for these findings. In the model, C and V inputs are filtered by parallel auditory streams that respond preferentially to transient and sustained properties of the acoustic signals, as in vision. These streams are represented by working memories that adjust their processing rates to cope with variable acoustic input rates. More rapid transient inputs can cause greater activation of the transient stream which, in turn, can automatically gain control the processing rate in the sustained stream. An invariant percept obtains when the relative activations of C and V representations in the two streams remain unchanged. The trading relation may be simulated as a result of how different experimental manipulations affect this ratio. It is suggested that the brain can use duration of a subsequent vowel to make the [b]/[w] distinction because the speech code is a resonant event that emerges between working memory activation patterns and the nodes that categorize them. [Article 3]

## **Central Pattern Generators**

Rhythmic processes are important in acoustical and speech perception and production. A class of central pattern generator models was developed to explain behavioral data about human and animal rhythmic gaits and their transitions. Simulated data include all the cat gait transitions (walk-trot-pace-gallop) originally studied by Pearson (1976), the human walk-run transition (Muybridge, 1957), and the transition from anti-phase movements to synchronous in-phase movements during increasingly rapid finger oscillations (Yamanishi *et al.*, 1980). These analyses clarified how a descending volitional, or GO, signal can interact with suitably designed cooperative-competitive feedback networks to generate only the desired gaits and transitions, quickly and reliably. They also clarified how switches from either anti-phase to in-phase, or in-phase to anti-phase movements can be generated as the GO signal increases in different parameter ranges of a single network. These transitions are due to bifurcations of the nonlinear system dynamics that are caused by the GO control parameter. Data about spinal cord and basal ganglia, among other brain structures, are interpreted in terms of the model. [Articles 17-19, 36, 41]

## **Dynamics of Nonlinear Oscillators: Vision, VLSI Circuit Design, and Motor Control**

Synchronous oscillations in visual cortex that generate coherent perceptual units have been reported in the neurophysiological literature (Gray and Singer, 1989). These oscillations appear to play a role in the binding of differing parts of a visual object into a coherent perceptual whole. This study modeled how neural networks could be designed to achieve rapid synchronized oscillations among distributed data even in high levels of cellular noise [Articles 29-32]. These networks also have been used to fit psychophysical data concerning temporal order judgments and threshold contrast as a function of stimulus size [Articles 32, 38], notably data of Hirsch and Sherrick (1961) and Essock (1990). It was also shown how stochastic resonance could arise in the model, thereby improving its signal-to-noise ratio in noise.

Cohen *et al.* [Article 39] has shown that a Metal Oxide Semiconductor Field Effect Transistor (MOSFET) circuit exhibits winner-take-all behavior. This circuit has been shown to be absolutely stable and is the basis for analog circuit design. This circuit shows promise as part of a larger VLSI architecture.

Cohen and Priebe [Article 22] have constructed dynamical equations using Clifford Algebras for the motion of the eyeball in the eye when governed by the pulley and no-pulley

models of motion. The no-pulley model assumes that rotation force is applied along three axes corresponding to the Euler Angle Coordinates. The pulley model assumes that force is applied tangent to the surface of the eyeball at the known point of insertion of eye muscles (Miller, 1989). Since the motion of the eye is inherently rotational, the torque applied to the eye is naturally described in the rotational notation provided by Clifford Algebras. Equations are specified by infinitesimal rotational velocities imparted along differing axes at different angular coordinates. These velocities can be used to provide control signals for training and control of a neural network [Article 40].

## ADDITIONAL REFERENCES

- Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Carpenter, G.A. and Grossberg, S. (Eds.) (1991). *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press.
- Carpenter, G.A. and Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, **16**, 131–137.
- Essock, E.A. (1990). The influence of stimulus length on the oblique effect of contrast sensitivity. *Vision Research*, **30**(8) 1243–1246.
- Gray, C.M. and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex, *Proceedings of the National Academy of Sciences USA*, **86**, 1698–1702.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **87**, 1–51.
- Halperin, L. (1977). The effect of harmonic ratio relationships on auditory stream segregation. Technical Report, McGill University, Psychology Department.
- Hirsch, I.J. and Sherrick, C.E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, **62**(5), 423–432.
- Miller, G.A. and Licklider, J.C.R. (1950). Intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, **22**, 167–173.
- Miller, J.M. (1989). Functional anatomy of the normal human rectus muscles. *Vision Research*, **29**(2), 223–240.
- Moore, B.C.J., Glasberg, B.R., and Peters, R.W. (1985). Relative dominance of individual partials in determining the pitch of complex tones *Journal of the Acoustical Society of America*, **77**, 1853–1860.
- Muybridge, E. (1957). *Animals in motion*. Dover Publishers (reprinted from the 1887 manuscript).
- Pearson, K.G (1976). The control of walking. *Scientific American*, **235**, 72–86.
- Repp, B.H. (1980). A range-frequency effect on perception of silence in speech. Haskins Laboratories Status Report on Speech Research, SR-61, pp. 151–165.
- Ritsma, R.J. and Engle, F.L. (1964). Pitch of frequency-modulated signals. *Journal of the Acoustical Society of America*, **36**, 1637–1644.
- Steiger, H. (1980). Some informal observations concerning the perceptual organization of patterns containing frequency glides. Tech Report, McGill University.
- Tougas, Y. and Bregman, A.S. (1990). The crossing of auditory streams. *Journal of Experimental Psychology: Human Perception and Performance*, **11**, 788–798.
- Yamanishi, J., Kawato, M., and Suzuki, R. (1980). Two coupled oscillators as a model for the coordinated finger tapping by both hands. *Biological Cybernetics*, **37**, 219–225.

## **SELECTED ABSTRACTS**

# **NEURAL NETWORK MODELS OF TEMPORAL PROCESSING IN SPEECH PERCEPTION AND MOTOR CONTROL**

Ian Boardman

PhD Dissertation, Boston University, 1994  
Department of Cognitive and Neural Systems

## **Abstract**

This dissertation studies three classes of neural networks for temporal processing: the first models storage of temporally ordered motor commands; the second, temporal integration of rapidly changing and quasi-static acoustic segments for rate-sensitive phonetic discrimination; and the third, rate-invariant grouping of phonetic percepts. Human performance in recalling planned movement sequences and spoken word lists from working memory indicates that output rate depends dynamically on the number of remaining items. In chapter 1, a neural network model is developed that represents temporally ordered events as spatial activation patterns. Events are performed at a rate that is regulated by an internal measure of memory load and an externally imposed arousal signal. Performance can be interrupted and restarted using arousal, and made dependent on efferent feedback. The model explains list length and position effects, and predicts that subjects tend toward a uniform output rate at moderate arousal. Variations in model parameters can account for variability in performance across subjects.

Chapter 2 models local speech rate effects on phonetic discrimination; in particular, how the duration of a subsequent vowel can shift the category boundary between recognition of stop and semi-vowel for the prior consonant. The model uses complementary processing of the transient and sustained components of the speech signal to generate phonetic decision contours that obey the observed exponential relation between formant transition rate and vowel duration.

Chapter 3 models adaptation of phonetic percepts to global speech rate as exemplified by shifts of category boundaries between single and double voiced stops due to changes in mean closure interval. A dynamical neural network model of a working memory and a list category field is developed which automatically adjusts its integration rate, or gain, according to the rate of input presentations. Computer simulations quantitatively generate the experimentally observed category boundaries for voiced stop pairs that have the same or different place of articulation. To explain why the closure interval required to hear a double (geminate) stop is typically more than twice as long as that needed to hear two different stops, the model depends upon resonant feedback between list categories and working memory. Resonance facilitates category formation and sustains expectation of the category until a reset occurs, either rapidly due to mismatch or slowly due to transmitter depletion.

## **VARIABLE RATE WORKING MEMORIES FOR PHONETIC CATEGORIZATION AND INVARIANT SPEECH PERCEPTION**

Ian Boardman, Michael Cohen, and Stephen Grossberg

Technical Report CAS/CNS-TR-93-008, Boston University  
In **Proceedings of the World Congress on Neural Networks**  
Hillsdale, NJ: Erlbaum Associates, 1993, **III**, pp. 2-5

### **Abstract**

Speech can be understood at widely varying production rates. A working memory is described for short-term storage of temporal lists of input items. The working memory is a cooperative-competitive neural network that automatically adjusts its integration rate, or gain, to generate a short-term memory code for a list that is independent of item presentation rate. Such an invariant working memory model is used to simulate data of Repp (1980) concerning the changes of phonetic category boundaries as a function of their presentation rate. Thus the variability of categorical boundaries can be traced to the temporal invariance of the working memory code.

---

Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225 and AFOSR 90-0128), ARPA (ONR N00014-92-J-4015), and the Office of Naval Research (ONR N00014-91-J-4100).

# **NEURAL DYNAMICS OF PHONETIC TRADING RELATIONS FOR VARIABLE-RATE CV SYLLABLES**

Ian Boardman†, Stephen Grossberg‡, and Michael Cohen§

Technical Report CAS/CNS-TR-94-037, Boston University

## **Abstract**

The perception of CV syllables exhibits a trading relationship between voice onset time (VOT) of a consonant and duration of a vowel. Percepts of [ba] and [wa] can, for example, depend on the durations of the consonant and vowel segments, with an increase in the duration of the subsequent vowel switching the percept of the preceding consonant from [w] to [b]. A neural model, called PHONET, is proposed to account for these findings. In the model, C and V inputs are filtered by parallel auditory streams that respond preferentially to transient and sustained properties of the acoustic signals, as in vision. These streams are represented by working memories that adjust their processing rates to cope with variable acoustic input rates. More rapid transient inputs can cause greater activation of the transient stream which, in turn, can automatically gain control the processing rate in the sustained stream. An invariant percept obtains when the relative activations of C and V representations in the two streams remain unchanged. The trading relation may be simulated as a result of how different experimental manipulations affect this ratio. It is suggested that the brain can use duration of a subsequent vowel to make the [b]/[w] distinction because the speech code is a resonant event that emerges between working memory activation patterns and the nodes that categorize them.

---

† Supported in part by the Advanced Research Projects Agency (AFOSR 90-0083), the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), and Pacific Sierra Research Corporation (PSR 91-6075-2).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

§ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

# STORE WORKING MEMORY NETWORKS FOR STORAGE AND RECALL OF ARBITRARY TEMPORAL SEQUENCES

Gary Bradskit†, Gail A. Carpenter‡, and Stephen Grossberg§

Technical Report CAS/CNS-TR-92-028, Boston University  
*Biological Cybernetics*, 1994, **71**, 469–480

## Abstract

Neural network models of working memory, called Sustained Temporal Order REcurrent (STORE) models, are described. They encode the invariant temporal order of sequential events in short term memory (STM) in a way that mimics cognitive data about working memory, including primacy, recency, and bowed order and error gradients. As new items are presented, the pattern of previously stored items is invariant in the sense that relative activations remain constant through time. This invariant temporal order code enables all possible groupings of sequential events to be stably learned and remembered in real time, even as new events perturb the system. Such a competence is needed to design self-organizing temporal recognition and planning systems in which any subsequence of events may need to be categorized in order to control and predict future behavior or external events. STORE models show how arbitrary event sequences may be invariantly stored, including repeated events. A preprocessor interacts with the working memory to represent event repeats in spatially separate locations. It is shown why at least two processing levels are needed to invariantly store events presented with arbitrary durations and interstimulus intervals. It is also shown how network parameters control the type and shape of primacy, recency, or bowed temporal order gradients that will be stored.

---

† Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0128) and the Office of Naval Research (ONR N00014-91-J-4100 and ONR N00014-92-J-1309).

‡ Supported in part by British Petroleum (BP 89A-1204), ARPA (AFOSR 90-0083 and ONR N00014-92-J-4015), the National Science Foundation (NSF IRI-90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).

§ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), ARPA (AFOSR 90-0083 and ONR N00014-92-J-4015) and the Office of Naval Research (ONR N00014-91-J-4100 and ONR N00014-92-J-1309).

# A PRELIMINARY LOOK AT A FAST LEARNING ARCHITECTURE FOR SPEAKER INDEPENDENT SPEECH RECOGNITION

Gary Bradski† and Michael Cohen‡

In **Proceedings of the World Congress on Neural Networks**  
Hillsdale, NJ: Erlbaum Associates, 1993 **III**, pp. 37–39

## Abstract

This paper extends a supervised, quick learning, generalized nearest neighbor network to learning and recognizing times series—in this case, speaker-independent speech waveforms. The supervised nearest neighbor network used here is fuzzy ARTMAP [2] which generalizes nearest neighbor classifiers by using nearest hyper-boxes with an adjustable maximum box size set by supervised feedback. Fuzzy ARTMAP consists of a fuzzy ART [1] input categorizer which makes associations to a supervised MAP field. Fuzzy ARTMAP is extended here by using many fuzzy ART input fields, one for each time slice, all of which connect to the same MAP field. Speech inputs are optimally aligned to learned patterns in the fuzzy ARTs using **Dynamic Time Warping (DTW)**. Here optimality is defined as the alignment that produces the highest fuzzy ART nodal choice scores. Aligned temporal inputs are then learned by the ART networks and associations are made to “the word” or “not the word” in the MAP field. During recognition, inputs are optimally aligned, but only inputs that select “the word” count towards the recognition score. This network has been termed **ARTMAP Temporal Extension Network with Dynamic time warping, or ATTEND**.

---

† Supported in part by the National Science Foundation (NSF IRI-90-24877) and the Office of Naval Research (ONR N00014-92-J-1309).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

# A NEURAL MODEL OF TIMED RESPONSE LEARNING IN THE CEREBELLUM

Daniel Bullock†, John C. Fiala‡, and Stephen Grossberg§

Technical Report CAS/CNS-TR-94-007, Boston University  
*Neural Networks*, 1994, 7, 1101-1114

## Abstract

A spectral timing model is developed to explain how the cerebellum learns adaptively timed responses during the rabbit's nictitating membrane response (NMR). The model posits two learning sites that respectively enable conditioned excitation and timed disinhibition of the response. Long-term potentiation of mossy fiber pathways projecting to interpositus nucleus cells allows conditioned excitation of the response's adaptive gain. Long-term depression of parallel fiber-Purkinje cell synapses in the cerebellar cortex allows learning of an adaptively timed reduction in Purkinje cell inhibition of the same nuclear cells. A spectrum of partially timed responses summate to generate an accurately timed population response. In agreement with physiological data, the model Purkinje cell activity decreases in the interval following the onset of the conditioned stimulus, and nuclear cell responses match conditioned response (CR) topography. The model reproduces key behavioral features of the NMR, including the properties that CR peak amplitude occurs at the unconditioned stimulus (US) onset; a discrete CR peak shift occurs with a change in interstimulus interval (ISI) between conditioned stimulus (CS) and US; mixed training at two different ISIs produces a double-peaked CR; CR acquisition and rate of responding depend unimodally on the ISI; CR onset latency decreases during training; and maladaptively timed, small-amplitude CRs result from ablation of cerebellar cortex.

---

† Supported in part by the National Science Foundation (NSF IRI-90-24877) and the Office of Naval Research (ONR N00014-92-J-1309).

‡ Supported in part by the Office of Naval Research (ONR N00014-92-J-1309).

§ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-92-J-1309).

# EVALUATION OF SPEAKER NORMALIZATION METHODS FOR VOWEL RECOGNITION USING FUZZY ARTMAP AND K-NN

Gail A. Carpenter† and Krishna K. Govindarajan‡

Technical Report CAS/CNS-TR-93-013, Boston University  
In **Proceedings of the World Congress on Neural Networks**  
Hillsdale, NJ: Erlbaum Associates, 1993, III, pp. 10-15

## Abstract

A procedure that uses fuzzy ARTMAP and K-Nearest Neighbor (K-NN) categorizers to evaluate intrinsic and extrinsic speaker normalization methods is described. Each classifier is trained on preprocessed, or normalized, vowel tokens from about 30% of the speakers of the Peterson-Barney database, then tested on data from the remaining speakers. Intrinsic normalization methods included one nonscaled, four psychophysical scales (bark, bark with end-correction, mel, ERB), and three log scales, each tested on four different combinations of the fundamental ( $F_0$ ) and the formants ( $F_1, F_2, F_3$ ). For each scale and frequency combination, four extrinsic speaker adaptation schemes were tested: centroid subtraction across all frequencies (CS), centroid subtraction for each frequency (CSi), linear scale (LS), and linear transformation (LT). A total of 32 intrinsic and 128 extrinsic methods were thus compared. Fuzzy ARTMAP and K-NN showed similar trends, with K-NN performing somewhat better and fuzzy ARTMAP requiring about 1/10 as much memory. The optimal intrinsic normalization method was bark scale, or bark with end-correction, using the differences between all frequencies (Diff All). The order of performance for the extrinsic methods was LT, CSi, LS, and CS, with fuzzy ARTMAP performing best using bark scale with Diff All; and K-NN choosing psychophysical measures for all except CSi.

---

† Supported in part by British Petroleum (BP 89A-1204), ARPA (AFOSR (90-0083 and ONR N00014-92-J-4015), the National Science Foundation (NSF IRI-90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), ARPA (ONR N00014-92-J-4015), and the National Science Foundation (NSF IRI-90-00530).

# **PARALLEL AUDITORY FILTERING BY SUSTAINED AND TRANSIENT CHANNELS SEPARATES COARTICULATED VOWELS AND CONSONANTS**

Michael A. Cohen† and Stephen Grossberg‡

Technical Report CAS/CNS-TR-93-051, Boston University

## **Abstract**

A neural model of peripheral auditory processing is described and used to separate features of coarticulated vowels and consonants. After preprocessing of speech via a filterbank, the model splits into two parallel channels, a sustained channel and a transient channel. The sustained channel is sensitive to relatively stable parts of the speech waveform, notably synchronous properties of the vocalic portion of the stimulus. It extends the dynamic range of eighth nerve filters using coincidence detectors that combine operations of raising to a power, rectification, delay, multiplication, time averaging, and preemphasis. The transient channel is sensitive to critical features at the onsets and offsets of speech segments. It is built up from fast excitatory neurons that are modulated by slow inhibitory interneurons. These units are combined over high frequency and low frequency ranges using operations of rectification, normalization, multiplicative gating, and opponent processing. Detectors sensitive to frication and to onset and offset of stop consonants and vowels are described. Model properties are characterized by mathematical analysis and computer simulations. Neural analogs of model cells in the cochlear nucleus and inferior colliculus are noted, as in an analog between the proposed sustained and transient auditory processing with sustained and transient visual processing.

---

† Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

‡ Supported in part by the Advanced Research Projects Agency (ONR N00014-92-J-4015) and the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

# NEURAL CONTROL OF INTERLIMB COORDINATION AND GAIT TIMING IN BIPEDS AND QUADRUPEDS

Michael A. Cohen†, Stephen Grossberg‡, and Christopher Priebe§

Technical Report CAS/CNS-TR-93-004, Boston University

Submitted to *Journal of Neurophysiology*

## Abstract

1) A family of central pattern generators, called GO Gait Generators, is described in which both the frequency and the relative phase of oscillations are controlled by a scalar arousal or GO signal that instantiates the will to act. The model cells obey shunting membrane equations, and interact via fast excitatory feedback signals and slow inhibitory feedback signals, organized as an on-center off-surround anatomy.

2) With two excitatory cells, or cell populations, the model describes an opponent processing network in which both in-phase and anti-phase oscillations can occur at different arousal levels. This two-channel oscillator can also produce phase transitions from either in-phase to anti-phase oscillations, or anti-phase to in-phase oscillations, in different parameter ranges, as the GO signal increases.

3) The two-channel oscillator is used to simulate data from human bimanual finger coordination tasks in which anti-phase oscillations at low frequencies spontaneously switch to in-phase oscillations at high frequencies, in-phase oscillations can be performed both at low and high frequencies, phase fluctuations occur at the anti-phase in-phase transition, and a "seagull effect" of larger errors occurs at intermediate phases. When driven by environmental patterns with intermediate phase relationships, the model's output exhibits a tendency to slip toward purely in-phase and anti-phase relationships as observed in humans subjects.

4) A four-channel oscillator is used to simulate quadruped vertebrate gaits, including the amble, the walk, all three pairwise gaits (trot, pace, and gallop), and the pronk. Spatial or temporal asymmetries in oscillator activation by the GO signal can trigger these transitions. Rapid transitions are simulated in the order—walk, trot, pace, and gallop—that occurs in the cat.

5) This precise switching control is achieved by using GO-dependent modulation of the model's inhibitory interactions that generates a different functional connectivity in a single network at different arousal levels. Such task-specific modulation of functional connectivity in neural pattern generators has been experimentally reported in invertebrates. A role for such a mechanism in gait-switching is predicted to occur in vertebrates.

6) A four channel oscillator can generate the two standard human gaits: the walk and the run. Although these two gaits are qualitatively different, they both have the same limb order and may exhibit oscillation frequencies that overlap. The model simulates the walk and the run via qualitatively different waveform shapes. The fraction of cycle that activity is above threshold quantitatively distinguishes the two gaits, much as the duty cycles of the feet are longer in the walk than in the run.

---

† Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0128 and AFOSR F49620-92-J-0225).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175 and AFOSR F49620-92-J-0225), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-92-J-1309).

§ Supported in part by the Army Research Office (ARO DAAL03-88-K-0088), the Advanced Research Projects Agency (AFOSR 90-0083), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-92-J-1309).

# A SPECTRAL NETWORK MODEL OF PITCH PERCEPTION

Michael A. Cohen†, Stephen Grossberg‡, and Lonce Wyse\*

Technical Report CAS/CNS-TR-92-024, Boston University  
*Journal of the Acoustical Society of America*, in press, 1995

## Abstract

A model of pitch perception, called the Spatial Pitch Network or SPINET model, is developed and analysed. The model neurally instantiates ideas from the spectral pitch modeling literature and joins them to basic neural network signal processing designs to simulate a broader range of perceptual pitch data than previous spectral models. The components of the model are interpreted as peripheral mechanical and neural processing stages, which are capable of being incorporated into a larger network architecture for separating multiple sound sources in the environment.

The core of the new model transforms a spectral representation of an acoustic source into a spatial distribution of pitch strengths. The SPINET model uses a weighted "harmonic sieve" whereby the strength of activation of a given pitch depends upon a weighted sum of narrow regions around the harmonics of the nominal pitch value, and higher harmonics contribute less to a pitch than lower ones. Suitably chosen harmonic weighting functions enable computer simulations of pitch perception data involving mistuned components, shifted harmonics, and various types of continuous spectra including rippled noise. It is shown how the weighting functions produce the dominance region, how they lead to octave shifts of pitch in response to ambiguous stimuli, and how they lead to a pitch region in response to the octave-spaced Shepard tone complexes and Deutsch tritones without the use of attentional mechanisms to limit pitch choices. An on-center off-surround network in the model helps to produce noise suppression, partial masking, and edge pitch. Finally, it is shown how peripheral filtering and short term energy measurements produce a model pitch estimate that is sensitive to certain component phase relationships.

---

† Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), ARPA (ONR N00014-92-J-4015), and the Office of Naval Research (ONR N00014-91-J-4100).

Supported in part by the American Society for Engineering Education, and the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

# **STATISTICAL PROPERTIES OF SINGLE AND COMPETING NONLINEAR FAST-SLOW OSCILLATORS IN NOISE**

Stephen Grossberg† and Alexander Grunewald‡

Technical Report CAS/CNS-TR-93-022, Boston University  
In **Proceedings of the World Congress on Neural Networks**  
Hillsdale, NJ: Erlbaum Associates, 1993, IV, 303-307

## **Abstract**

Statistical properties of fast-slow Elias-Grossberg oscillators are studied in response to deterministic and noisy inputs. Oscillatory responses remain stable in noise due to the slow inhibitory variable, which establishes an adaptation level that centers the oscillatory responses of the fast excitatory variable to deterministic and noisy inputs. Competitive interactions between oscillators improve the stability in noise. Although individual oscillation amplitudes decrease with input amplitude, the average total activity increases with input amplitude, thereby suggesting that oscillator output is evaluated by a slow process at downstream network sites.

---

† Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), ARPA (ONR N00014-92-J-4015), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-91-J-4100).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

# NEURAL CONTROL OF INTERLIMB OSCILLATIONS, I: HUMAN BIMANUAL COORDINATION

Stephen Grossberg<sup>†</sup>, Christopher Priebe<sup>§</sup>, and Michael A. Cohen<sup>†</sup>

Technical Report CAS/CNS-TR-94-021, Boston University  
Submitted to *Biological Cybernetics*

## Abstract

How do humans and other animals accomplish coordinated movements? How are novel combinations of limb joints rapidly assembled into new behavioral units that move together in in-phase or anti-phase movement patterns during complex movement tasks? A neural central pattern generator (CPG) model simulates data from human bimanual coordination tasks. As in the data, anti-phase oscillations at low frequencies switch to in-phase oscillations at high frequencies, in-phase oscillations occur both at low and high frequencies, phase fluctuations occur at the anti-phase in-phase transition, a "seagull effect" of larger errors occurs at intermediate phases, and oscillations slip toward in-phase and anti-phase when driven at intermediate phases. These oscillations and bifurcations are emergent properties of the CPG model in response to volitional inputs. The CPG model is a version of the Elias-Grossberg oscillator. Its neurons obey Hodgkin-Huxley type equations whose excitatory signals operate on a faster time scale than their inhibitory signals in a recurrent on-center off-surround anatomy. When an equal command or GO signal activates both model channels, the model CPG can generate both in-phase and anti-phase oscillations at different GO amplitudes. Phase transitions from either in-phase to anti-phase oscillations, or from anti-phase to in-phase oscillations, can occur in different parameter ranges, as the GO signal increases.

---

<sup>†</sup> Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0499 and AFOSR F49620-92-J-0225), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-92-J-1309).

<sup>§</sup> Supported in part by the Army Research Office (ARO DAAL03-88-K-0088), the Advanced Research Projects Agency (AFOSR 90-0083), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-92-J-1309).

<sup>†</sup> Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0128 and AFOSR F49620-92-J-0225).

**BINDING OF OBJECT REPRESENTATIONS  
BY SYNCHRONOUS CORTICAL DYNAMICS EXPLAINS  
TEMPORAL ORDER AND SPATIAL POOLING DATA**

Alexander Grunewald and Stephen Grossberg†

In **Proceedings of the 16th Annual Conference  
of the Cognitive Science Society**

A. Ram and K. Eiselt (Eds.)

Hillsdale, NJ: Erlbaum Associates, 1994, pp. 387-391

**Abstract**

A key problem in cognitive science concerns how the brain binds together parts of an object into a coherent visual object representation. One difficulty that this binding process needs to overcome is that different parts of an object may be processed by the brain at different rates and may thus become desynchronized. Perceptual framing is a mechanism that resynchronizes cortical activities corresponding to the same retinal object. A neural network model based on cooperation between oscillators via feedback from a subsequent processing stage is presented that is able to rapidly resynchronize desynchronized featural activities. Model properties help to explain perceptual framing data, including psychophysical data about temporal order judgments. These cooperative model interactions also simulate data concerning the reduction of threshold contrast as a function of stimulus length. The model hereby provides a unified explanation of temporal order and threshold contrast data as manifestations of a cortical binding process that can rapidly resynchronize image parts which belong together in visual object representations.

---

† Supported in part by the Advanced Research Projects Agency (ONR N00014-92-J-4015) and the Office of Naval Research (ONR N00014-91-J-4100).

# **INVARIANT SPEECH RECOGNITION AND AUDITORY OBJECT FORMATION: NEURAL MODELS AND PSYCHOPHYSICS**

Krishna K. Govindarajan

PhD Dissertation, Boston University  
Department of Cognitive and Neural Systems, 1994

## **Abstract**

This dissertation investigates three topics concerning variability and robustness in speech perception: variability of the speech signal across speakers, variability due to speaking rate effects, and the robustness of speech perception in noisy environments.

Given that the speech signal corresponding to a given phoneme can vary considerably across speakers, invariant speech perception can be facilitated by normalizing the signal across speakers. In chapter 2, 160 intrinsic and extrinsic speaker normalization methods are compared using a neural network, fuzzy ARTMAP, and K-Nearest Neighbor (K-NN) categorizers trained and tested on disjoint sets of speakers of the Peterson-Barney vowel database. ARTMAP and K-NN show similar trends, with K-NN performing better but requiring about ten times as much memory. The optimal intrinsic normalization method is bark scale using the differences between all frequencies, while the optimal extrinsic method is linear transformation of the vowel space to a canonical representation.

In chapter 3, psychophysical studies of adaptation to the mean silence duration between two different stop consonants are examined. Using natural speech stimuli, the first experiment shows that the category boundary between hearing only one or hearing both stop consonants varied as a function of the distribution of silent intervals. The second experiment shows that the variance of the distribution did not significantly affect the boundary, and the final experiment shows sequential effects in the adaptation process. Finally, a model of the adaptation process is developed which emulates the data.

In environments with multiple sound sources, the auditory system is capable of teasing apart the impinging jumbled signal into different mental objects. Chapter 4 presents a neural network model of auditory scene analysis, which groups different frequency components based on pitch and spatial location cues and allocates the components to different objects. While location primes the grouping mechanism, segregation is based solely on harmonicity. The model qualitatively emulates results from psychophysical grouping experiments, such as how a tone sweeping upwards in frequency groups due to frequency proximity with a downward sweeping tone even if noise exists at the intersection point; and illusory percepts, such as the illusion of a tone continuing through noise.

## INFLUENCE OF SILENCE DURATION DISTRIBUTION IN PERCEPTION OF STOP CONSONANT CLUSTERS

Krishna K. Govindarajan† and Michael A. Cohen‡

*Journal of the Acoustical Society of America*, 1994, **95**, 2978

### Abstract

The effects of the silence duration between two different stop consonants in /VC<sub>1</sub>-C<sub>2</sub>V/ tokens were examined. Earlier results using synthetic stimuli had shown that the decision boundary between hearing one and two stop consonants varied as a function of the range and frequency (number of tokens) of the silent interval [B.H. Repp, Haskins Lab Status Report on Speech Research, **SR-61**, 151-165 (1980)]. Using real speech stimuli, the Repp (1980) experiment was replicated, showing that the results hold for individual subjects. Extending this result, the overall variance of the silence duration distribution was manipulated while maintaining the same mean, for two different distributions. The result shows that the overall variance does not significantly affect the boundary, and that listeners must be adapting to the mean silence duration. The final experiment examines this adaptation by presenting trials of sequential tokens to determine the relative weighting of the former token(s) on the judgment of the final token in the trial. The results of this experiment will be presented in relation to the adaptation process. Implications for adaptation in other durational cues, and influence of cue trading will be discussed.

---

† Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), British Petroleum (BP 89-1204), and the National Science Foundation (NSF IRI-90-00530).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

# A NEURAL NETWORK MODEL OF AUDITORY SCENE ANALYSIS AND SOURCE SEGREGATION

Krishna K. Govindarajan†, Stephen Grossberg‡, Lonce L. Wyse§, and Michael A. Cohen¶

Technical Report CAS/CNS-TR-94-039, Boston University

## Abstract

In environments with multiple sound sources, the auditory system is capable of teasing apart the impinging jumbled signal into different mental objects, or streams, as in its ability to solve the cocktail party problem. A neural network model of auditory scene analysis, called the ARTSTREAM model, is presented that groups different frequency components based on pitch and spatial location cues, and selectively allocates the components to different streams. The grouping is accomplished through a resonance that develops between a given object's pitch, its harmonic spectral components, and (to a lesser extent) its spatial location. Those spectral components that are not reinforced by being matched with the top-down prototype read-out by the selected object's pitch representation are suppressed, thereby allowing another stream to capture these components, as in the "old-plus-new heuristic" of Bregman. These resonance and matching mechanisms are specialized versions of Adaptive Resonance Theory, or ART, mechanisms. The model is used to simulate data from psychophysical grouping experiments, such as how a tone sweeping upwards in frequency creates a bounce percept by grouping with a downward sweeping tone due to proximity in frequency, even if noise replaces the tones at their intersection point. The model also simulates illusory auditory percepts such as the auditory continuity illusion of a tone continuing through a noise burst even if the tone is not present during the noise, and the scale illusion of Deutsch whereby downward and upward scales presented alternately to the two ears are regrouped based on frequency proximity, leading to a bounce percept. The stream resonances provide the coherence that allows one voice or instrument to be tracked through a multiple source environment.

---

† Supported in part by the Advanced Research Projects Agency (ONR N00014-92-J-4015), the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), British Petroleum (BP 89A-1204), and the National Science Foundation (NSF IRI-90-00530).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

§ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225) and by the American Society for Engineering Education.

¶ Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

## SUFFICIENT CONDITIONS FOR WTA BEHAVIOR OF A SHUNTING INHIBITION NETWORK

T.G. Kincaid and M.A. Cohen†

Submitted for publication to *Neural Networks*

### Abstract

In this paper, three classes of artificial neural networks, which exhibit Winner-Take-All (WTA) behaviour are studied. Sufficient conditions for the convergence of these networks to the WTA equilibrium points are presented. The properties of resetting or non resetting when the external inputs change for WTA behavior are investigated in some detail.

---

† Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225).

**NEURAL DYNAMICS OF GAZE AND GAIT:  
SPATIAL SENSORY-MOTOR CONTROL AND  
TEMPORAL PATTERN GENERATION**

Christopher A. Priebe

PhD Dissertation, Boston University, 1994  
Department of Cognitive and Neural Systems

**Abstract**

Neural networks together with their emergent properties can provide a mechanistic basis for explaining quantifiable regularities in human and animal behavior. In this dissertation, neural bases of observed spatial and temporal regularities are separately analysed: Part I treats the control of eye movements in 3-D and Part II treats the generation of rhythmic skeletomotor patterns.

a neural network theory of oculomotor control is advanced that explains regularities of saccadic eye movements in space, namely Listing's and Donders' Laws. The theory is tested with the aid of a new biomechanical model of how the eye is moved in its orbit by the six extra-ocular muscles. Listing's Law is a regularity found in eye movements that has puzzled students of human behavior for over a century. The set of eye rotations that obey this law are a subset of the much larger set of rotations that can be realized by the oculomotor plant. The theory shows how eye rotations that obey Listing's Law may emerge from a reactive saccadic control system that translates visual eccentricity into motor commands via a fixed spatial gradient. In such an untrained system, targets can be foveated through successive approximation. The eye attitudes assumed upon successful foveation provide a statistically stable basis for learning that improves the accuracy of first saccades. Simulation of an adaptive saccadic control system shows that foveation-gated, error-based learning in motor coordinates improves saccade accuracy without compromising adherence to Listing's Law.

A family of central pattern generators defined by cooperative-competitive nonlinear networks is developed to parametrically fit many data about gait timing and transitions. A scalar arousal or GO signal drives the oscillations and controls their frequency, relative phase, and waveform shape. A two-channel model is used to simulate data from human bimanual finger coordination tasks. Four-channel models simulate properties of quadruped and biped gaits, including the cat walk-trot-pace-gallop and human walk-run sequences. GO-dependent modulation of connection strengths controls fast transitions in simulated quadruped gaits.

## **NEURAL CONTROL OF INTERLIMB OSCILLATIONS, II: BIPED AND QUADRUPED GAITS AND BIFURCATIONS**

Christopher Priebe§, Stephen Grossberg†, and Michael A. Cohen‡

Technical Report CAS/CNS-TR-94-022, Boston University

Submitted to *Biological Cybernetics*

### **Abstract**

Behavioral data concerning animal and human gaits and gait transitions are simulated as emergent properties of a central pattern generator (CPG) model. The CPG model is a version of the Elias-Grossberg oscillator. Its neurons obey Hodgkin-Huxley type equations whose excitatory signals operate on a faster time scale than their inhibitory signals in a recurrent on-center off-surround anatomy. A descending command or GO signal activates the gaits and triggers gait transitions as its amplitude increases. A single model CPG can generate both in-phase and anti-phase oscillations at different GO amplitudes. Phase transitions from either in-phase to anti-phase oscillations, or from anti-phase to in-phase oscillations, can occur in different parameter ranges, as the GO signal increases. Quadruped vertebrate gaits, including the amble, the walk, all three pairwise gaits (trot, pace, and gallop), and the pronk are simulated using this property. Rapid gait transitions are simulated in the order—walk, trot, pace, and gallop—that occurs in the cat, along with the observed increase in oscillation frequency. Precise control of quadruped gait switching uses GO-dependent modulation of inhibitory interactions, which generates a different functional anatomy at different arousal levels. The primary human gaits (the walk and the run) and elephant gaits (the amble and the walk) are simulated by oscillations with the same phase relationships but different waveform shapes at different GO signal levels, much as the duty cycles of the feet are longer in the walk than in the run. Relevant neural data from spinal cord, globus pallidus, and motor cortex, among other structures, are discussed.

---

§ Supported in part by the Army Research Office (ARO DAAL03-88-K-0088), the Advanced Research Projects Agency (AFOSR 90-0083), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-92-J-1309).

† Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0499 and AFOSR F49620-92-J-0225), the National Science Foundation (NSF IRI-90-24877), and the Office of Naval Research (ONR N00014-92-J-1309).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0128 and AFOSR F49620-92-J-0225).